

# 基于链接实体回放的多源知识图谱终身表示学习\*

孙泽群, 崔员宁, 胡伟



(计算机软件新技术国家重点实验室(南京大学), 江苏 南京 210023)

通信作者: 胡伟, E-mail: whu@nju.edu.cn

**摘要:** 知识图谱存储大量的结构化知识和丰富的语义信息, 已被广泛应用于知识驱动的智能软件. 随着智能应用的不断发展, 它们对知识图谱的需求也在发生变化. 而单一知识图谱往往具有数据不完备等缺点, 难以满足需求. 因此, 支持新数据来源、融合多源知识已成为迫切需求. 传统的知识图谱表示学习和应用范式只考虑单一图谱, 忽视了不同图谱间的知识迁移. 多源知识图谱联合训练虽然可以带来性能提升, 但不支持新增知识图谱的拓展表示学习. 鉴于此, 提出了多源知识图谱终身表示学习的新范式. 给定一个知识图谱序列, 终身表示学习的目标是在学习新知识图谱的同时, 从已学习的知识图谱与模型中获得知识迁移. 为实现这一目标, 提出了一个基于链接实体回放的多源知识图谱终身表示学习框架. 首先, 设计了一个以 Transformer 为编码器的知识图谱表示学习模型作为框架核心, 利用关系相关性进行实体的链接预测; 其次, 提出了链接子图构造方法, 基于实体对齐构建并回放新增知识图谱和已有知识图谱之间的链接子图进行终身学习和知识迁移; 最后, 采用动态结构方法, 为每个知识图谱存储相应的模型参数快照来避免灾难性遗忘. 多个链接预测基准数据集上的实验结果表明: 所提出的表示学习模型可以取得最先进的性能, 且提出的终身表示学习框架可以实现有效的知识迁移.

**关键词:** 多源知识图谱; 表示学习; 终身学习; 回放; 链接预测

**中图法分类号:** TP18

中文引用格式: 孙泽群, 崔员宁, 胡伟. 基于链接实体回放的多源知识图谱终身表示学习. 软件学报, 2023, 34(10): 4501-4517. <http://www.jos.org.cn/1000-9825/6887.htm>

英文引用格式: Sun ZQ, Cui YN, Hu W. Lifelong Representation Learning of Multi-sourced Knowledge Graphs via Linked Entity Replay. Ruan Jian Xue Bao/Journal of Software, 2023, 34(10): 4501-4517 (in Chinese). <http://www.jos.org.cn/1000-9825/6887.htm>

## Lifelong Representation Learning of Multi-sourced Knowledge Graphs via Linked Entity Replay

SUN Ze-Qun, CUI Yuan-Ning, HU Wei

(State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210023, China)

**Abstract:** Knowledge graphs (KGs) store a great amount of structured knowledge and semantic information. They have been widely used by many knowledge-powered intelligent applications. With the rapid development of these applications, their requirements for knowledge also change. A single KG usually suffers from the incompleteness issue and is therefore unable to meet the requirement. This suggests an urgent demand for supporting new data sources and fusing multi-sourced knowledge. The conventional paradigm for KG representation learning and application only considers a single KG while ignores the knowledge transfer between different sources. Joint representation learning on multi-sourced KGs can bring performance improvement, but it cannot support the extended representation learning of new KGs. To resolve these issues, this paper presents a new paradigm, i.e., lifelong representation learning on multi-sourced KGs. Given a sequence of multi-sourced KGs, lifelong representation learning aims at benefiting from the previously-learned KG and embedding model when learning a new KG. To this end, this study proposes a lifelong learning framework based on linked entity replay. First, it designs a

\* 基金项目: 国家自然科学基金(62272219)

本文由“知识赋能的信息系统”专题特约编辑高宏教授、陈华钧教授、赵翔教授、李瑞轩教授推荐.

收稿时间: 2022-07-05; 修改时间: 2022-08-18, 2022-12-14; 采用时间: 2022-12-28; jos 在线出版时间: 2023-01-13

Transformer-based KG embedding model that leverages relation correlations for link prediction between entities. Second, it proposes a linked subgraph generation method. It leverages the entity alignment between different sources to build the subgraph and replays the linked entities to enable lifelong learning and knowledge transfer. Finally, it uses a dynamic model structure with model parameters and embeddings stored for each KG to avoid catastrophic forgetting. Experiments on benchmarks show that the proposed KG embedding model can achieve the state-of-the-art performance in link prediction, and the lifelong representation learning framework is effective and efficient in multi-sourced knowledge transfer compared with baselines.

**Key words:** multi-sourced knowledge graph; representation learning; lifelong learning; replay; link prediction

近年来,知识图谱(knowledge graph, KG)凭借结构化的知识表示与丰富的语义信息,逐渐成为人工智能技术的重要组成部分,为智能化应用如语义搜索、智能问答、个性化推荐等提供知识来源<sup>[1]</sup>.知识图谱表示学习是支持这些智能应用的重要研究方向,它将知识图谱的符号化表示转化为向量表示,学习潜在语义信息,方便下游应用进行语义计算<sup>[2,3]</sup>.然而,知识图谱往往具有领域单一、知识不完备与稀疏等缺点,无法满足智能应用日益增长的知识需求.与单一知识图谱相比,多源图谱拥有的互补知识可以更好地满足下游应用的需求.然而,下游应用对新知识图谱与领域知识的需求往往是逐步出现的,难以事先获取所有知识图谱.因此,真实的应用场景需要一种新的知识图谱表示学习范式来支持不断增长与变化的知识需求.

传统的知识图谱表示学习与应用范式如图 1(a)所示:随着多源知识图谱的不断到来,需要在每个新知识图谱上重新训练一个表示学习模型.在这一范式中,知识图谱模型的训练相互独立,没有利用图谱之间的互补性知识.与之不同的是,近期的相关研究表明:基于实体对齐进行多源知识图谱联合训练<sup>[4,5]</sup>,如图 1(b)所示,可以有效学习更具表达力的知识图谱向量表示,进而提高下游的任务性能.但联合训练需要事先获取所有知识图谱,不适用于下游任务所需知识源不断更新和迭代的场景.为了解决这些问题,本文提出了多源知识图谱终身表示学习.如图 1(c)所示,随着任务的不断迭代,不同的知识图谱也逐渐加入.终身表示学习(lifelong representation learning)的目标是:在进行新知识图谱学习的同时,复用从先前知识图谱中学习到的语义信息,促进知识迁移,从而提高任务性能.

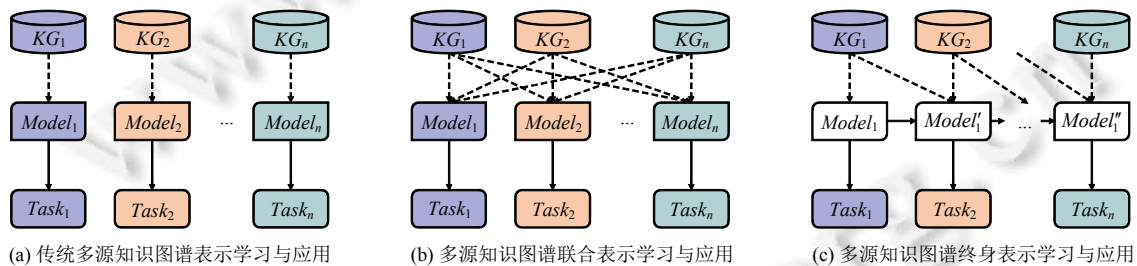


图 1 多源知识图谱表示学习与应用范式对比

多源知识图谱终身表示学习存在以下难点:其一,如何设计高效且可迁移的知识图谱表示学习模型,这是进行终身表示学习的基础;其二,如何将已学习的知识迁移到新增知识图谱的表示学习模型中,这是多源知识图谱上实现终身表示学习的关键,现有的独立训练和联合训练方法都无法做到这一点;其三,如何在终身学习的同时防止参数覆盖,避免灾难性遗忘.针对上述3个挑战,本文提出一种结合链接实体回放与动态结构的终身表示学习框架 LifeKE,它主要包含3项技术:一个基于关系相关性聚合邻居信息的 Transformer 编码器、一种基于链接子图筛选与回放的高效知识迁移方法和一个基于动态结构的模型扩展方式.复杂度分析与实验结果表明:LifeKE 不仅在单图谱上具有较强的表示学习能力,而且在新增的知识图谱上也能够高效复用已学习知识指导新模型学习,同时还能够以很低的参数成本保留已学习的知识.本文的主要贡献概括如下:

- (1) 提出了多源知识图谱终身表示学习新范式,充分利用多源知识图谱之间的互补信息进行知识迁移.相较于传统的表示学习和联合表示学习,所提范式可以提高多源知识图谱表示学习的效果和效率;
- (2) 提出了一个基于链接实体回放的多源知识图谱终身表示学习框架 LifeKE,它基于关系相关性进行

链接预测来充分捕捉关系语义, 并使用链接实体回放的方法迁移知识和降低迁移成本;

- (3) 在知识图谱链接预测任务上开展实验, 结果表明: 所提模型在多个基准数据集上取得了目前最先进的性能, 多源知识图谱终身表示学习可有效提高任务性能, 且主要技术模块均有助于性能提升.

本文第 1 节介绍知识图谱表示学习和终身学习的相关工作. 第 2 节介绍问题定义和整体框架. 第 3 节详述模型与方法. 第 4 节进行实验评估与结果分析. 第 5 节总结全文并展望未来研究方向.

## 1 相关工作

本文的相关工作包括知识图谱表示学习和终身学习.

### 1.1 知识图谱表示学习

知识图谱表示学习旨在为实体和关系学习低维向量表示, 在向量空间中计算原始符号空间中的语义信息. 现有工作大致可分为 3 类.

- 第 1 类是基于几何运算的模型, 例如 TransE<sup>[6]</sup>、TransH<sup>[7]</sup>、TransR<sup>[8]</sup>、TransD<sup>[9]</sup>、TransAH<sup>[10]</sup>和 RotatE<sup>[11]</sup>等. 这类方法通过向量平移、旋转等运算构建和解释实体与关系的语义信息. 例如, 给定三元组  $(s, r, o)$ , 其中,  $s$  表示主语实体,  $r$  表示关系,  $o$  表示宾语实体, TransE 将关系解释为实体间的平移向量, 即满足  $\mathbf{s} + \mathbf{r} = \mathbf{o}$ , 其中, 粗体表示对应元素的向量表示;
- 第 2 类是语义匹配模型, 其使用基于相似性的评分函数度量语义信息. 代表性工作包括 DistMult<sup>[12]</sup>、Complex<sup>[13]</sup>、Simple<sup>[14]</sup>和 TuckER<sup>[15]</sup>等. 例如, DistMult 利用双线性函数对三元组进行打分;
- 最后一类是采用深度神经网络的模型, 通过端到端的方式预测实体与关系的潜在语义信息. 一般来说, 这类方法的训练目标是: 给定主语实体和关系, 通过神经网络预测对应的宾语实体. 代表性工作有基于卷积神经网络的 ConvE<sup>[16]</sup>、基于循环跳跃网络的 RSN<sup>[17]</sup>、基于图卷积网络的 CompGCN<sup>[18]</sup>和 M<sup>2</sup>GNN<sup>[19]</sup>以及基于 Transformer<sup>[21]</sup>的 CoKE<sup>[20]</sup>等. 以 CoKE 为例, 给定三元组  $(s, r, o)$ , 其随机遮蔽掉一个实体, 如  $(s, r, [mask])$ , 然后使用 Transformer<sup>[21]</sup>进行表示学习, 并预测被遮蔽的实体.

此外, 还有相关工作结合知识图谱中的显式知识和语言模型中的隐式事实进行表示学习<sup>[22]</sup>. 读者可以参考相关综述<sup>[2,3,23]</sup>对知识图谱表示学习进行全面了解.

相较于上述面向单知识图谱的表示学习模型, 多源知识图谱表示学习的工作主要关注实体对齐. 这些工作设计了不同的实体对齐目标, 如向量空间映射<sup>[24]</sup>、校准<sup>[25]</sup>与交互<sup>[26]</sup>等. 为了进一步增强实体对齐的效果, 一些工作引入实体的多视图<sup>[27-32]</sup>或多模态<sup>[33,34]</sup>特征, 如实体名称、属性与高阶邻居子图等. 另外一些工作考虑半监督学习, 如自训练(self-training)和协同训练(co-training)等迭代式方法<sup>[26,35,36]</sup>, 解决实体对齐训练数据不足的问题. 读者可以参考相关综述了解基于表示学习的实体对齐的相关进展<sup>[37,38]</sup>. 需要注意的是, 本文工作不是面向实体对齐任务, 而是基于实体对齐进行多源知识图谱的终身表示学习和知识迁移, 目标是提高知识图谱的表示学习效果, 进而增强下游任务性能. 此外, 本文工作与动态图谱或时序图谱等工作也不同. 动态图谱考虑单个图谱的增删改等动态变化; 时序图谱考虑事实的时间信息; 本文工作考虑的是多个不同图谱.

### 1.2 终身学习

终身学习(lifelong learning)也被称为持续学习或连续学习(continual learning), 旨在按照一定的顺序学习连续非同分布的流数据, 增量式地学习新任务和更新模型参数<sup>[39]</sup>. 与传统的机器学习范式不同, 终身学习的数据资源和学习任务是持续增加的. 这就要求机器学习模型不仅能够保留已学习的旧知识, 还要具有继续学习新知识的能力. 直接采用梯度更新方法对模型进行增量更新, 会使模型在已学习任务上的性能迅速下降, “忘记”已学到的知识, 这一问题被称为灾难性遗忘(catastrophic forgetting)问题, 是终身学习研究领域的核心问题. 针对这一问题, 目前主要有 3 类解决方法: 基于正则化的方法, 如 EWC<sup>[40]</sup>, 通过建模计算各参数对已学习任务的重要性, 在学习新知识时对各参数施以正则化约束, 防止重要参数被遗忘; 基于记忆回放的方法,

如 GEM<sup>[41]</sup>, 选取和保存部分已学习任务数据, 在学习新任务时回放已学习任务数据, 使模型能够保持在已学习任务上的性能; 基于动态结构的方法, 如 DEN<sup>[42]</sup>, 在每个新增任务上增加新参数, 在不干扰之前任务的基础上学习新任务. 本文提出的 LifeKE 是一种结合记忆回放与动态结构的方法, 基于动态结构保持旧任务上的性能, 并在此基础上采用选择性记忆回放增强新任务的学习. 与传统的基于回放的方法不同, 由于动态结构已经能够避免对旧任务的遗忘, 本文所采用的记忆回放侧重迁移旧知识来增强新任务的学习, 而非维持在旧任务上的性能.

传统的终身学习相关研究大多针对图像分类、关系抽取等分类任务开展. 近年来, 随着知识图谱工程应用的推进, 对于可扩展的知识图谱表示学习的需求日益增长, 知识图谱的终身表示学习逐渐受到关注. 文献 [43] 较早提出终身表示学习问题, 并尝试结合多种经典的终身学习方法来实现知识图谱的终身表示学习. DiCGRL<sup>[44]</sup> 基于关系类型将知识图谱分解为多个子图, 当新增事实加入知识图谱时, 仅更新与该事实最相关的子图, 并通过回放该事实的邻居来避免灾难性遗忘. ABLE<sup>[45]</sup> 基于预设锚点来稳定知识图谱的表示空间, 在初次训练时, 基于已有图谱选择中心锚点, 在学习新事实时更新和回放锚点最短路径上的事实, 从而保持表示空间的稳定. 另外, GraIL<sup>[46]</sup>、MEAN<sup>[47]</sup>、LAN<sup>[48]</sup> 等方法在初始知识图谱上学习如何归纳式地表示实体, 它们在新数据上不再增量式地更新模型, 而是直接归纳式地生成新增图谱的表示. 尽管上述几种方法可以持续学习单一知识图谱的新增事实, 但它们都依赖单知识图谱稳定的模式结构, 因此不具备本文讨论的多源知识图谱终身表示学习的能力. 如果将某知识图谱更新后的版本看作一个“新图谱”, 与旧图谱构成多源知识图谱, 则本文的技术路线也可以用来处理这一场景. 相关探索留作未来工作. 此外, 本文提出的 LifeKE 也可以看作一种增量模型, 基于已经学到的模型, 增加新参数(而非更新参数)来学习新知识图谱并进行知识迁移.

本文还与图或异质信息网络的预训练工作相关<sup>[49,50]</sup>. 该类工作首先基于自监督学习预训练模型, 然后通过参数微调的方式应用于下游任务. 除了所处理的数据不同外, 它们与本文工作的主要区别在于知识迁移的场景不同. 预训练相关工作基于大数据和自监督任务训练大模型, 进一步挖掘和抽象数据特征, 增强下游模型与任务. 其目标是特征的深度抽取, 并以参数微调来进行知识迁移. 本文工作是在两个不同的模型之间, 通过局部数据回放和表示对齐来增强新模型的训练, 旧模型不进行微调. 未来工作可以考虑开展多源知识图谱预训练.

## 2 多源知识图谱终身表示学习框架

本节首先介绍背景知识并给出问题定义, 然后概述所提出的多源知识图谱终身表示学习框架.

### 2.1 背景与问题定义

本文将知识图谱  $\mathcal{K}$  定义为一个由实体、关系和三元组事实组成的知识库, 即  $\mathcal{K}=(\mathcal{E}, \mathcal{R}, \mathcal{T})$ . 其中,  $\mathcal{E}$  与  $\mathcal{R}$  分别表示实体和关系集合,  $\mathcal{T} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$  表示三元组集合. 每个三元组描述一个事实, 如  $\langle \text{Tim Berners-Lee}, \text{Awards}, \text{Turing Award (2016)} \rangle$  描述了蒂姆·伯纳斯-李获得了 2016 年图灵奖. 目前, 很多开源知识图谱被广泛应用, 如 Freebase<sup>[51]</sup>、DBpedia<sup>[52]</sup>、Wikidata<sup>[53]</sup> 和 YAGO<sup>[54]</sup>, 对推动知识驱动的智能应用的发展产生了重要影响. 为了促进知识融合, 链接开放数据(linked open data, LOD)项目通过发布和链接 Web 中的各类资源, 在不同知识源之间创建语义关联. 实体对齐是 LOD 的核心.

实体对齐描述了不同数据源中的共指实体(即指向真实世界中的同一对象). 给定两个不同的知识图谱  $\mathcal{K}_1$  和  $\mathcal{K}_2$ , 其实体对齐  $\mathcal{A}_{1,2}$  定义为  $\mathcal{A}_{1,2} = \{(e_1, e_2) \in \mathcal{E}_1 \times \mathcal{E}_2 | e_1 \equiv e_2\}$ , 其中,  $\equiv$  表示等价关系. 例如, 英文维基百科中的“Tim Berners-Lee”和中文百度百科的“蒂姆·伯纳斯-李”就是等价实体, 二者表示真实世界中的同一个人. 实体对齐为多源知识图谱搭建了知识迁移的桥梁. 本文提出的多源知识图谱终身表示学习也基于实体对齐.

在给出本文研究的问题定义之前, 先定义知识图谱表示学习. 给定知识图谱  $\mathcal{K}$ , 表示学习模型为其中的每一个实体和关系学习对应的  $d$  维向量. 本文将该模型定义为  $M(\mathcal{K}, \Theta, \mathbf{E})$ , 其中,  $\Theta$  表示模型自身参数集合, 如深度神经网络的权重矩阵和偏移向量等;  $\mathbf{E}$  表示知识图谱的实体和关系向量集合.  $\Theta$  和  $\mathbf{E}$  为需要学习的参数. 通

常情况下, 模型自身参数  $\Theta$  的规模远小于知识图谱向量表示  $E$  的规模. 例如: TransE 没有模型参数, 仅通过实体和关系向量的几何运算来捕捉语义信息; CompGCN 的模型参数是每层图神经网络的权重矩阵等; CoKE 的模型参数是 Transformer 里面的权重矩阵等. 这些模型参数的数量都远远小于实体和关系向量的参数量. 在本文研究的终身表示学习场景下, 多源知识图谱往往是独立构建的, 异构性较强, 这使得本文研究的重点与传统终身学习不同. 其他领域的终身学习主要目的是避免模型更新后遗忘旧知识. 而在本文场景下, 一方面, 模型参数较少; 另一方面, 异构性导致模型参数较难复用. 故本文的研究目标是致力于在学习新知识图谱的向量表示的同时, 以最小的代价从旧知识图谱和模型中获得收益.

本文研究多源知识图谱终身表示学习问题, 对于给定的多源知识图谱序列  $(\mathcal{K}_1, \mathcal{K}_2, \dots, \mathcal{K}_N)$ , 终身表示学习为每个知识图谱学习实体和关系的向量表示及其对应模型, 即  $M_1(\mathcal{K}_1, \Theta_1, E_1), M_2(\mathcal{K}_2, \Theta_2, E_2), \dots, M_N(\mathcal{K}_N, \Theta_N, E_N)$ , 使得  $M_i$  可以从  $M_{i-1}$  中获得知识迁移 ( $i > 1$ ), 进而间接地从更早的知识图谱和模型中获益. 本文假设序列中相邻的两个知识图谱具有实体对齐, 即  $|A_{i-1, i}| > 0$ , 但是不假设任意两个图谱直接存在实体对齐. 对于开源知识图谱而言, 实体对齐可以从 LOD 项目直接获得, 或者根据等价关系的传递性间接获得. 如何获取实体对齐不是本文的研究内容. 本文的研究重点是高效地表示学习模型和终身表示学习方法.

### 2.2 多源知识图谱终身表示学习框架

如图 2 所示, 本文所提框架将相邻知识图谱间的对齐实体作为知识迁移的桥梁, 在多源知识图谱上逐个学习新增知识图谱. 框架主要包含两个技术模块: 一个基于关系相关性的知识图谱表示学习模型和一种多源知识图谱终身表示学习方法. 基于关系相关性的表示学习模型利用实体已有关系和目标关系的相关性聚合实体信息, 捕捉知识图谱中的关系语义. 该模型是学习具有表现力的向量表示的关键, 也是进行终身学习和知识迁移的基础. 终身表示学习方法旨在持续地学习多源知识图谱的表示, 因为新知识图谱存在大量新实体和新关系, 所以终身表示学习的重点是学习新实体和新关系的向量表示, 而常用的对原模型增量更新的方式<sup>[40]</sup>无法表示新实体和新关系. 为了增强新向量表示的学习, 本文方法基于数据迁移和参数迁移进行终身表示学习, 其中, 可迁移数据是指包含对齐实体的三元组, 可迁移参数是指对齐实体的向量表示. 在终身学习的过程中, 所提框架保存每个知识图谱的实体以及关系向量表示和对应的表示学习模型. 由于传统的单图谱训练和多图谱联合训练也需要保存每个实体和关系的向量表示, 因此它们与本文框架所需存储空间几乎相同(具体参数复杂度分析见第 3.3 节).

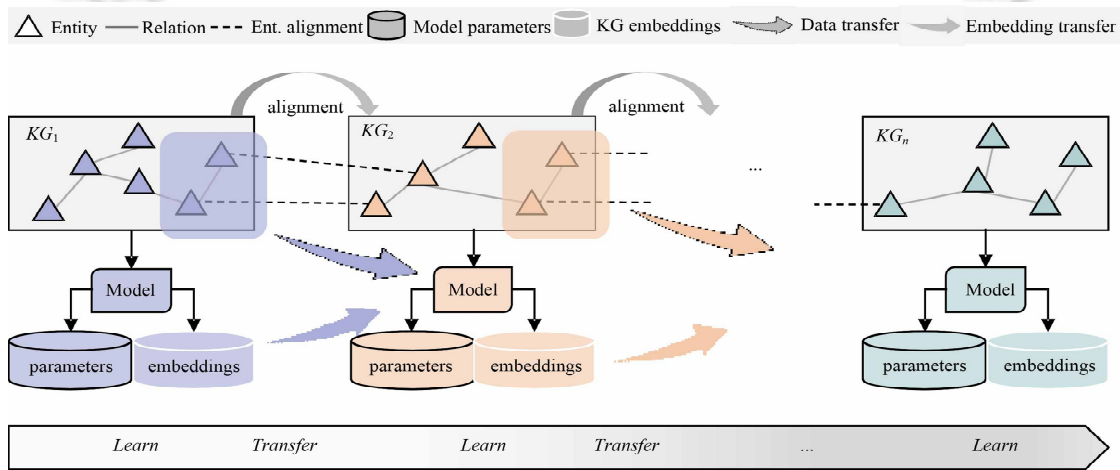


图 2 多源知识图谱终身表示学习框架

## 3 LifeKE 方法

本节介绍多源知识图谱终身表示学习框架下的具体模型与方法.

### 3.1 基于关系相关性的知识图谱表示学习

知识图谱表示学习的重要任务是捕捉实体间的关系语义,使得学到的向量表示和模型具有知识补充的能力,从而缓解知识图谱不完备的问题.给定三元组 $(s,r,o)$ ,本文借鉴 Transformer<sup>[20,21]</sup>的掩码预测方法遮蔽掉其中一个实体,比如宾语实体  $t$ ,构造模型输入 $(s,r,[mask]_o)$ ,经过 Transformer 网络,可以获得对应元素的表示:

$$v_s, v_r, v_o = \text{Encoder}(s, r, [mask]_o) \tag{1}$$

其中,斜粗体字母表示对应元素的输入向量表示,  $v$  表示编码器的输出表示.  $v_o$  表示替代了实体  $o$  的掩码的输出表示.本文用  $\text{Encoder}(\cdot)$  表示 Transformer 网络,作为三元组编码器.其核心组件是自注意力机制,使得输入的每一个元素均参与其他元素的注意力计算,进而捕捉任意元素间的重要性.这里不展开介绍技术细节,读者可以参考 Transformer 原文了解具体计算方式.  $[mask]_o$  作为 Transformer 的输出表示,捕捉到了来自主语实体  $s$  和关系  $r$  的信息,可以用来预测目标实体,即  $o$ .假定候选目标是图谱中的全部实体  $\mathcal{E}$ ,模型通过计算  $[mask]_o$  与实体向量的内积及  $\text{softmax}$  值来获取预测的概率分布  $p_o$ .例如,概率分布中预测目标实体是  $o$  的概率值为

$$p([mask] = o | (s, r, [mask]_o)) = \frac{\exp(v_o \cdot o)}{\sum_{e \in \mathcal{E}} \exp(v_o \cdot e)} \tag{2}$$

上述预测方法没有显式地利用主语实体  $s$  的已有信息,也就是实体的邻居三元组信息.为了增强实体表示和链接预测,本文提出一种基于关系相关性的注意力机制来以不同权重聚合邻居实体信息.

#### 3.1.1 基于关系相关性的注意力机制

链接预测旨在根据已知事实推理某关系的未知宾语实体,不同关系的三元组对目标关系的链接预测具有不同的重要性.例如:要预测某个研究生的研究方向,该研究生的就读学院和导师等信息要比学生的出生地更加重要.基于此,本文根据关系相关度来聚合这些关系邻居信息.关系相关度定义为多个关系经常共现来描述实体的概率.举例来说,“父亲”和“母亲”经常共现来描述人物信息,二者具有很强的相关性.这一定义类似现有工作中的属性相关性<sup>[25]</sup>.具体而言,本节提出一种基于关系相关性的注意力机制来增强实体表示,基于向量内积来计算关系相关性.如图 3 所示:对于知识图谱中实体  $s$  的  $n$  个入边关系的邻居三元组 $(e_1, r_1, s)$ ,  $(e_2, r_2, s), \dots, (e_n, r_n, s)$ ,首先根据惯例加入反向三元组,使入边三元组实际包含了实体  $s$  在原知识图谱里面的所有出入边信息;然后,基于关系的共现来计算注意力权重,例如,关系  $r_1$  与链接预测中的关系  $r$  的相关性定义为  $v_{r_1} \cdot r$ ,其中,  $v_{r_1}$  表示  $r_1$  经过编码器的输出表示,  $r$  表示关系  $r$  的初始化表示.基于此,本文将实体  $e$  的已有关系与链接预测目标关系的注意力定义为

$$w_{r_1, r} = \frac{v_{r_1} \cdot r}{\sum_{i=1}^n v_{r_i} \cdot r} \tag{3}$$

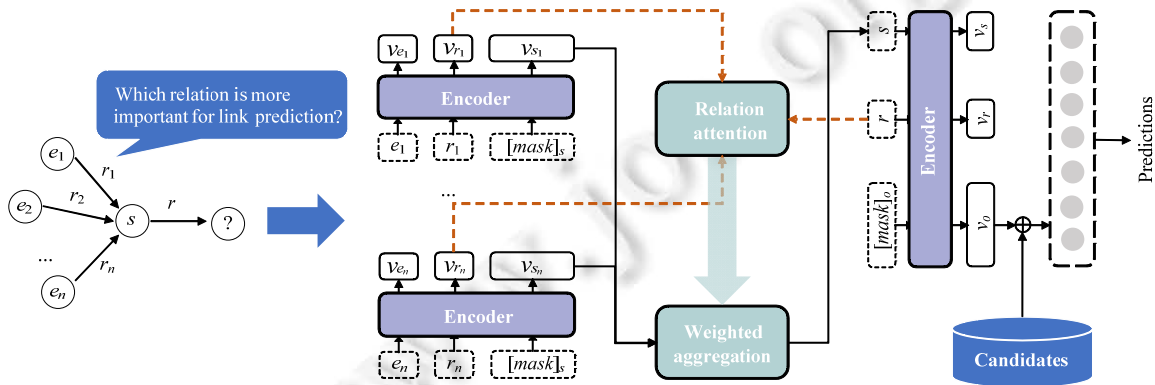


图 3 基于关系相关性的知识图谱表示学习模型

基于计算得到的注意力值,模型可以对相关的邻居信息进行带权聚合,进而增强实体  $e$  的表示.为此,需要将这些邻居三元组 $(e_1, r_1, s), (e_2, r_2, s), \dots, (e_n, r_n, s)$ 输入编码器.类似于公式(1),首先遮蔽掉每个邻居三元组中的

实体  $s$ , 构造如  $(e_1, r_1, [mask]_s)$  的输入. 然后, 将编码器输出的  $mask$  位表示  $\mathbf{v}_{s_1}$  作为实体  $s$  在当前邻居三元组中的表示. 注意, 尽管这里的邻居编码器和上文的三元组编码器网络结构相同, 但它们是两个不同的编码器, 具有不同的随机初始参数. 最后, 利用关系注意力值作为权重聚合来自不同邻居的实体表示:

$$\mathbf{s}' = \sum_{i=1}^n w_{r_i, r} \mathbf{v}_{s_i} \quad (4)$$

其中,  $\mathbf{s}'$  表示实体  $s$  来自邻居信息的表示, 它和随机初始化的向量表示  $\mathbf{s}$  平均后作为三元组编码器的输入, 即  $\mathbf{v}_s, \mathbf{v}_r, \mathbf{v}_o = \text{Encoder}\left(\frac{1}{2}(\mathbf{s}' + \mathbf{s}), \mathbf{r}, [mask]_o\right)$ . 这样, 本文提出的模型可以有选择地从实体的邻居信息中获得增益.

### 3.1.2 模型优化

本文的模型优化目标分为两部分. 考虑到长尾实体邻居信息少等问题, 除了上节介绍的基于邻居信息聚合之外, 本文还保留了不聚合邻居信息的实体表示作为三元组编码器的输入. 将实体  $s$  的两个表示(基于邻居信息聚合的表示和自身随机初始化的表示)分别输入到三元组编码器, 获得的输出表示分别用于预测概率计算, 得到两个预测概率分布:  $\mathbf{p}_o$  (基于自身表示) 和  $\mathbf{p}_o^n$  (基于邻居信息聚合).

令  $\mathbf{l}_o$  表示标签分布的独热(one-hot)向量, 即目标实体位置上是 1, 其余位置为 0. 本文使用交叉熵损失函数  $CE(\cdot)$  计算预测概率和真实标签之间的损失. 考虑到知识图谱有大量多映射关系, 如一对多关系(即一对主语实体和关系可能会有多个正确的宾语实体), 独热向量的标签概率分布过于绝对. 参考近期工作<sup>[20]</sup>, 本文考虑使用软标签进行模型训练. 具体来说, 给定概率值  $\alpha$ , 原标签向量  $\mathbf{l}_o$  中的 1 改成  $\alpha$ , 其余值改为  $\frac{1-\alpha}{|\mathcal{E}|-1}$ , 其中,  $|\mathcal{E}|$  表示候选实体的数量, 也是标签  $\mathbf{l}_o$  的维度. 这些软标签有利于学习候选集中存在多个正确宾语实体的训练样本.

本文设置知识图谱表示学习的损失函数如下:

$$\mathcal{L}_{\text{age}} = \frac{1}{|\mathcal{T}|_{(s,r,o) \in \mathcal{T}}} \sum CE(\mathbf{p}_o, \mathbf{l}_o) + \frac{1}{|\mathcal{T}|_{(s,r,o) \in \mathcal{T}}} \sum CE(\mathbf{p}_o^n, \mathbf{l}_o) \quad (5)$$

由于已经为知识图谱添加反向关系和事实, 原知识图谱的主语实体预测和宾语实体预测可以统一为现图谱的宾语实体预测. 加入反向边可以保证每一个实体均有入边和出边, 方便统一建模.

值得注意的是, 本工作与现有基于图神经网络的知识图谱表示学习模型<sup>[55]</sup>不同. 现有的基于图神经网络的知识图谱表示学习模型, 如 CompGCN, 主要分为两部分: 一是利用图神经网络聚合邻居信息来表示实体, 类似本文的邻居编码器; 二是利用几何模型(如 TransE)或语义匹配模型(如 DistMult)的能量函数或得分函数来评估三元组. 这两部分相互独立. 而本文模型是一个端到端的预测模型, 邻居信息的聚合也考虑了邻居关系和预测目标关系的相关性.

## 3.2 基于链接实体回放的终身表示学习

本小节介绍所提出的终身表示学习方法. 其主要挑战在于: 面对具有不同模式结构的多源知识图谱, 无法简单地对已有图谱的模型和向量进行微调来实现知识的迁移. 故本文借鉴实体对齐领域的空间对齐方法, 构造不同图谱之间的链接子图来迁移旧图谱的知识, 增强新图谱的训练. 具体来说, 方法包括链接子图采样和基于回放的空间对齐两个模块.

### 3.2.1 基于关系共现的链接子图采样

对于已经训练的现有知识图谱  $\mathcal{K}_{i-1}$  和新增的知识图谱  $\mathcal{K}_i$  以及它们的实体对齐  $\mathcal{A}_{i-1, i}$ , 令  $\mathcal{E}_{i-1, i}$  表示链接实体集合(即  $\mathcal{A}_{i-1, i}$  中出现过的实体, 也就是存在对齐关系的实体). 根据相关工作<sup>[26]</sup>, 链接子图定义为包含链接实体的三元组集合, 即  $\mathcal{G}_{i-1, i} = \{(s, r, o) \in \mathcal{T}_{i-1} | s \in \mathcal{E}_{i-1, i} \text{ or } o \in \mathcal{E}_{i-1, i}\} \cup \{(s, r, o) \in \mathcal{T}_i | s \in \mathcal{E}_{i-1, i} \text{ or } o \in \mathcal{E}_{i-1, i}\}$ , 其中,  $\mathcal{T}_i$  表示  $\mathcal{K}_i$  的三元组集合. 为融合链接子图中的共指实体, 可以让它们互换在各自三元组中的角色, 构造跨知识图谱的三元组. 举例来说, 给定链接子图中的两个共指实体  $(e_{i-1}, e_i)$  及其三元组  $(e_{i-1}, r_{i-1}, o_{i-1})$  与  $(e_i, r_i, o_i)$ , 其中,  $e_{i-1} \in \mathcal{E}_{i-1}$ ,  $(e_{i-1}, r_{i-1}, o_{i-1}) \in \mathcal{T}_{i-1}$ ,  $e_i \in \mathcal{E}_i$ ,  $(e_i, r_i, o_i) \in \mathcal{T}_i$ , 通过交换  $e_{i-1}$  和  $e_i$ , 可以构造两个新的跨知识图谱三元组, 即  $(e_i, r_{i-1}, o_{i-1})$

与 $(e_{i-1}, r_i, o_i)$ , 作为两个知识图谱的桥梁. 联合训练多个知识图谱及其链接子图, 既可捕捉到实体对齐信息, 使得共指实体具有相似的向量表示, 又可以进行跨图谱知识迁移, 缓解语义异构性问题. 目前的链接子图考虑了链接实体的所有一阶邻居信息. 由于一阶邻居的数量会远大于链接实体数量, 且不是所有的三元组都对新增知识图谱的表示学习有重要贡献, 因此本文改进了链接子图的构造方法来采样部分子图.

本文的链接子图采样算法基于关系共现频次选择部分子图, 其动机和上文的基于关系相关性的注意力机制异曲同工. 区别在于: 采样算法考虑跨图谱的关系相关性, 而非单一图谱内部的关系相关性. 为了实现这一目标, 本文将上述全量链接子图中的共指实体合并为一个虚拟实体节点, 使得该实体的关系包含了来自两个不同图谱的数据. 链接子图的采样过程如算法 1 所示, 首先获取合并的链接子图统计关系共现频次, 然后计算链接子图中所涉及的已有知识图谱关系和新增知识图谱关系的频次并降序排列, 最后保留前百分之  $k$  的旧关系, 其余旧关系及其对应的三元组则从链接子图中删除. 该链接子图筛选算法旨在减少回放过程中参与的实体数量, 减少重复训练的旧知识图谱实体的数量. 其中, 基于关系共现的筛选方法不仅能够明显减少实体和三元组数量, 还能够保证知识迁移的增益效果. 复杂度方面, 算法 1 主要包括共现频率统计和链接子图筛选两个步骤, 总时间复杂度为  $O(|\mathcal{A}_{i-1,i}|^2 + N_{g_{i-1,i}})$ , 其中,  $N_{g_{i-1,i}}$  为  $\mathcal{G}_{i-1,i}$  中三元组的数量.

**算法 1.** 基于关系共现的链接子图采样算法(LinkedSubgraphSampling).

输入:  $\mathcal{K}_{i-1}$  和  $\mathcal{K}_i$  以及实体对齐  $\mathcal{A}_{i-1,i}$ ;

输出: 链接子图的三元组集合  $\mathcal{P}$ .

```

1   $\mathcal{P} \leftarrow \text{Initialize}(\cdot)$  //空集合
2   $\mathcal{T}_{i-1}, \mathcal{T}_i \leftarrow \text{Add reverse triplets to } \mathcal{T}_{i-1}, \mathcal{T}_i$  //添加反向三元组
3   $\mathcal{G}_{i-1,i} \leftarrow \text{Retrieve one-hop relational neighbors of linked entities}$ ; //构造初始的链接子图
4  Compute co-occurrence frequency of relations in  $\mathcal{G}_{i-1,i}$ ; //统计关系共现频率
5   $\mathcal{R}'_{i-1} \leftarrow \text{Retrieve the most frequent } k\% \text{ old relations}$ ; //筛选出和新知识图谱相关的  $k\%$  旧关系
6  for  $(s,r,o) \in \mathcal{G}_{i-1,i}$  do //筛选链接子图
7    if  $r \in \mathcal{R}'_{i-1}$  do
8       $\mathcal{P}.add((s,r,o))$ ;
9    end if
10 end for
11 return  $\mathcal{P}$ 

```

### 3.2.2 基于链接子图回放的表示空间对齐

链接子图将和新知识图谱一起用于模型训练, 其中既包含旧知识图谱的部分实体、关系和三元组, 也包含新知识图谱的部分知识. 对齐实体是已学习知识图谱和新增知识图谱之间的桥梁, 现有实体对齐的相关工作表明: 在相同或对齐的语义空间中联合训练链接子图中的三元组, 能够利用实体对齐关系, 使得共指实体具有相同或相似的向量表示<sup>[26]</sup>. 与上述工作的同步联合训练不同, 本文面向知识图谱不断新增的终身学习场景, 各知识图谱的训练需要异步进行. 为此, 本文设计了一种异步的表示空间对齐方法, 实现从已学习模型到新增知识图谱的知识迁移. 具体来说, 首先将链接子图中的对齐实体作为锚点, 然后最小化这些实体的新旧向量表示的欧式距离, 对齐新图谱空间和旧图谱空间, 使得新增图谱的表示空间和模型间接学习到旧图谱的知识, 本文将这一过程称为链接实体的回放. 令  $\mathcal{E}'_{i-1}$  表示链接子图中的旧知识图谱  $\mathcal{K}_{i-1}$  实体集合, 则回放的损失函数定义为

$$\mathcal{L}_{\text{replay}} = \frac{1}{|\mathcal{E}'_{i-1}|} \sum_{e \in \mathcal{E}'_{i-1}} \|e_{i-1} - e_i\| \quad (6)$$

其中,  $e_{i-1}$  表示实体  $e$  在旧知识图谱  $\mathcal{K}_{i-1}$  训练时获得的向量, 而  $e_i$  表示其在链接子图与新知识图谱  $\mathcal{K}_i$  联合训练时的向量.  $\|\cdot\|$  计算向量的  $L_2$  范数. 相关工作表明<sup>[56]</sup>, 更新训练好的向量表示会造成不可避免的知识遗忘. 因此在本文工作中, 实体的新表示  $e_i$  不被用于更新其旧表示  $e_{i-1}$ . 这样, 旧知识图谱的知识被迁移到了新模型



和表示, 而旧表示不受影响.

为避免参数覆盖带来的灾难性遗忘, LifeKE 基于动态结构方法为每个知识图谱保存一个模型和一组实体与关系的向量表示. 尽管这种方法在传统的终身学习中会带来一定的模型存储成本, 但保存预训练好的知识图谱表示在实际应用中是非常有必要的. 在下面第 3.3 节的复杂度分析中, 可以看到: 在多源知识图谱终身表示学习场景下, LifeKE 保持了与知识图谱独立训练几乎相同的模型存储成本.

### 3.3 多源知识图谱终身表示学习训练流程

本节介绍多源知识图谱终身表示学习的训练流程, 如算法 2 所示. 对于学习序列中的第 1 个知识图谱, 使用第 3.1 节介绍的表示学习方法学习图谱中实体和关系的向量表示, 获得第 1 个知识图谱表示模型  $M_1(\mathcal{K}_1, \Theta_1, E_1)$ . 当新增知识图谱  $\mathcal{K}_i$  加入终身学习序列后, 首先基于算法 1 采样  $\mathcal{K}_i$  与旧知识图谱  $\mathcal{K}_{i-1}$  的局部链接子图, 然后将其合并到  $\mathcal{K}_i$  的三元组集合作为训练数据, 接下来随机初始化模型参数和新知识图谱的向量表示并执行训练过程, 此时训练的总损失将变为表示学习和回放损失的加权和. 当模型性能在下游任务的验证集上开始连续下降时终止训练, 最后保存模型参数和知识图谱向量表示.

**算法 2.** 多源知识图谱终身表示学习训练流程.

输入: 新知识图谱  $\mathcal{K}_i$ ; 已学习的旧知识图谱  $\mathcal{K}_{i-1}$ ; 新增的对齐实体集  $\mathcal{A}_{i-1,i}$ ; 已学习知识图谱模型  $M_{i-1}$ ; 训练批次大小  $bs$ ; 学习率  $lr$ ; 最大训练轮数  $T$ ;

输出: 知识图谱表示模型  $M_i$ .

- 1  $\mathcal{G}_{i-1,i} \leftarrow \text{LinkedSubgraphSampling}(\mathcal{K}_i, \mathcal{K}_{i-1}, \mathcal{A}_{i-1,i});$  //算法 1
- 2  $\mathcal{T}_i \leftarrow \mathcal{T}_i \cup \mathcal{G}_{i-1,i};$  //合并链接子图到新知识图谱
- 3  $\Theta_i \leftarrow \text{Randomly initialize trainable parameters};$  //初始化模型参数
- 4  $E_i \leftarrow \text{Randomly initialize trainable KG embeddings};$  //初始化图谱向量
- 5  $S = \text{ceil}\left(\frac{|\mathcal{T}_i|}{bs}\right);$  //计算每轮的训练步数
- 6 **for**  $i \leftarrow 1$  **to**  $T$  **do**
- 7     **for**  $j \leftarrow 1$  **to**  $S$  **do**
- 8         Feed the randomly sampled batch  $b \subset \mathcal{T}_i$  into  $M_i(\mathcal{K}_i, \Theta_i, E_i)$ ;
- 9         Compute the KG embedding loss  $\mathcal{L}_{kge}$ ; //计算表示学习模型损失
- 10         Compute the replay loss  $\mathcal{L}_{replay}$ ; //计算回放损失
- 11          $\mathcal{L}_{final} \leftarrow \mathcal{L}_{kge} + \alpha \mathcal{L}_{replay};$  //  $\alpha$  是一个超参数用于平衡
- 12          $\Theta_i \leftarrow \Theta_i - lr \nabla_{\Theta_i} \mathcal{L}_{final};$  //梯度更新
- 13          $E_i \leftarrow E_i - lr \nabla_{E_i} \mathcal{L}_{final};$  //梯度更新
- 14     **end for**
- 15     Terminate if the performance drops on the validation data; //终止训练
- 16 **end for**
- 17 Save  $\Theta_i$  and  $E_i$ ; //保存模型参数副本和知识图谱向量表示
- 18 **return**  $M_i$

下面分析独立训练、联合重训练和所提出框架 LifeKE 的参数与时间复杂度. 首先分析参数复杂度. 第 3.1 节所涉及的 Transformer 编码器的参数量远小于实体和关系的向量表示, 因此 LifeKE 框架与独立训练和联合重训练的参数复杂度均为  $O(N_e \times d + N_r \times d)$ , 其中,  $N_e$  和  $N_r$  分别为所有知识图谱上实体数目和关系数目,  $d$  为表示向量的维度. 然后分析时间复杂度, 在迭代训练的每个轮次中, 时间复杂度与训练数据数量成正比. 由于多来源的知识图谱是依次出现的, 因此联合重训练方法需要在每个知识图谱出现时重新联合训练所有图谱, 在批次大小不变的情况下, 其累计时间复杂度为  $O(N \times N_T)$ , 其中,  $N$  表示知识图谱数量,  $N_T$  为多源知识图谱中

累积的事实数量. 独立训练的时间复杂度为  $O(N_T)$ . LifeKE 的时间复杂度为  $O(N_T + \varepsilon)$ , 其中,  $\varepsilon$  为相关链接子图中用于迁移的事实数量, 通常远小于  $N_T$ . 综上所述, 多源知识图谱联合训练有较高的时间复杂度, 而本文提出的 LifeKE 与传统的独立训练方法有着几乎相同的参数复杂度和时间复杂度, 需要的额外资源很少.

## 4 实验

本节使用链接预测基准数据集来评估所提出的知识图谱表示学习模型和多源知识图谱终身表示学习框架的性能. 相关代码和数据在 GitHub 仓库 <https://github.com/nju-websoft/LifeKE> 开源.

### 4.1 链接预测

链接预测旨在预测不完备三元组中缺失的实体, 即给定  $(s, r, ?)$ , 预测缺失的宾语实体  $o$ . 链接预测是评估知识图谱表示学习性能最常用的任务.

#### 4.1.1 数据集介绍

FB15K-237<sup>[57]</sup>、WN18RR<sup>[16]</sup>和 YGAO3-10<sup>[54]</sup>是最常用的知识图谱链接预测数据集. FB15K-237 是 Freebase 知识图谱的子集, 其前身 FB15K<sup>[6]</sup>的训练集和测试集包含了大量反向事实, 如  $(s, r, o)$  和  $(o, r^{-1}, s)$ . 考虑到模型捕捉这种反向关系比较容易, FB15K-237 删除了这些反向事实. 因此, FB15K-237 比 FB15K 更具挑战性. WN18RR 是基于 WordNet<sup>[58]</sup>构建的链接预测数据集, 与 Freebase 包含的开放域知识不同, WordNet 是一个英语词典, 包含单词的词性词义和上下位关系等. 因此, WN18RR 的关系数量远少于 FB15K-237, 二者的领域知识差异较大. YGAO3-10 也是一个开放域知识图谱, 实体数量比 FB15K-237 多很多, 但是关系数量却很少. 这 3 个数据集的统计信息见表 1.

表 1 链接预测数据集统计

Dataset	#Entities	#Relations	#Training triples	#Valid. triples	#Test triples
FB15K-237	14 541	237	272 115	17 535	20 466
WN18RR	40 943	11	86 835	3 034	3 134
YAGO3-10	123 182	37	1 079 040	5 000	5 000

#### 4.1.2 实验设置

本文遵循链接预测任务的实验惯例. 给定一个测试三元组  $(s, r, o)$ , 构造两个测试用例  $(?, r, o)$  与  $(s, r, ?)$ , 分别预测主语和宾语实体, 并报告两种预测任务的平均性能. 指标方面, 本文报告常用的  $H@k$  ( $k=1, 3, 10$ ) 和 MRR (mean reciprocal rank). 对于每个测试用例, 模型会预测一个候选实体的排序列表.  $H@k$  计算正确候选实体排名在前  $k$  的测试用例的占比. MRR 表示候选实体排名倒数的均值.  $H@k$  和 MRR 数值越大, 表明模型的链接预测性能越好. 模型优化方面, 参数初始化方法采用 Xavier<sup>[59]</sup>, 优化器采用 Adam 算法<sup>[60]</sup>, 模型训练终止条件是在验证集上的  $H@1$  结果连续下降. 超参数设定基于网格搜索. 具体超参数设定可以参考 GitHub 代码仓库的运行脚本. 对比方法方面, 本文选取了目前最具代表性的 8 个知识图谱表示学习模型作为对比, 分别是 TransE、ComplEx、ConvE、RotatE、CompGCN、TuckER、CoKE 和 M<sup>2</sup>GNN. 可以参考第 2 节关于这些方法的介绍. 为保证公平对比, 本文直接采用这些对比方法在所用数据集上的官方结果. 对于在部分数据集上缺失官方结果的方法, 如 TransE 和 ComplEx, 本文工作使用其开源代码在缺失结果的数据集上进行复现.

#### 4.1.3 主要实验结果

表 2 给出了链接预测的结果对比, 其中, 最优结果用粗体表示, 次优结果用下划线表示. 可以看到, 本文提出的 LifeKE 模型在 FB15K-237 和 WN18RR 这两个最常用的数据集上均取得了先进的性能. 例如: 在数据集 FB15K-237 上, LifeKE 的  $H@1$  指标比 CoKE 提高了 2.2%, 比 M<sup>2</sup>GNN 提高了 1.1%. 所提模型性能提升的主要原因在于 Transformer 强大的表示学习能力和本文提出的基于关系相关性的注意力机制(请参考第 4.1.4 节关于注意力机制的消融实验). 其他对比方法中, 同样基于 Transformer 的 CoKE 表现最好, 但是由于它仅建模了三元组信息, 没有显式地利用实体的邻居信息和关系相关性, 所以其性能弱于本文提出的 LifeKE. 综合 LifeKE 与 CoKE 的表现可以看出: 基于深度学习尤其是 Transformer 的模型, 在知识图谱表示学习任务上具有

很大的潜力. 另外, 本文发现: 在 YAGO3-10 这一数据集上, 所提模型的效果尽管相较于 CoKE 提升很多, 却落后于 M<sup>2</sup>GNN. 这是由于 YAGO3-10 数据规模很大, 但是关系数量很少, 关系模式较为简单, 本文模型的关系邻居聚合难以发挥作用. 而 M<sup>2</sup>GNN 将图神经网络推广到混合曲率的表示空间, 可以自适应地建模不同难度和维度的关系模式, 在 YAGO3-10 数据集上更具优势.

表 2 链接预测结果比较

Models	FB15K-237				WN18RR				YAGO3-10			
	H@10	H@3	H@1	MRR	H@10	H@3	H@1	MRR	H@10	H@3	H@1	MRR
TransE	0.528	0.367	0.231	0.329	0.528	0.400	0.013	0.223	0.679	0.571	0.404	0.504
ComplEx	0.415	0.274	0.156	0.243	0.478	0.427	0.343	0.394	0.450	0.245	0.063	0.191
ConvE	0.501	0.356	0.237	0.325	0.520	0.440	0.400	0.430	0.620	0.490	0.350	0.440
RotatE	0.533	0.375	0.241	0.338	0.571	0.492	0.428	0.476	0.670	0.550	0.402	0.495
CompGCN	0.535	0.390	0.264	0.355	0.546	0.494	0.443	0.479	0.527	0.374	0.241	0.338
TuckER	0.544	0.394	0.266	0.358	0.526	0.482	0.443	0.470	0.647	0.517	0.380	0.472
CoKE	0.549	0.400	0.272	0.364	0.553	0.496	<b>0.450</b>	0.484	0.614	0.486	0.354	0.443
M <sup>2</sup> GNN	<b>0.565</b>	0.398	0.275	0.362	<b>0.572</b>	0.498	0.444	0.485	<b>0.702</b>	<b>0.605</b>	<b>0.478</b>	<b>0.543</b>
LifeKE	0.556	<b>0.408</b>	<b>0.278</b>	<b>0.371</b>	0.563	<b>0.500</b>	<b>0.450</b>	<b>0.487</b>	0.684	0.562	0.415	0.508

4.1.4 消融实验

为了验证所提出的基于关系相关性的注意力机制的有效性以及邻居信息对于链接预测的作用, 本节进行消融实验. 表 3 报告了所提模型分别去掉邻居信息聚合(即去掉  $\mathcal{L}_{age}$  中的第 2 部分)和关系注意力机制(即邻居信息平均聚合)的结果. 可以看到: 去除邻居信息后, 模型性能下降明显. 例如: 在 FB15K-237 上, LifeKE w/o neigh. 的 H@1 只有 0.270, 低于使用邻居信息聚合的结果 0.278, 这说明, 邻居信息可以帮助链接预测. 去除注意力机制后, LifeKE w/o attn. 在 FB15K-237、WN18RR 和 YAGO3-10 测试集的 H@1 结果分别是 0.275、0.441 和 0.411, 分别低于带有注意力机制的 0.278、0.450 和 0.415. 这些结果说明了关系注意力机制的有效性.

表 3 关于邻居信息和关系注意力机制的消融实验

Models	FB15K-237				WN18RR				YAGO3-10			
	H@10	H@3	H@1	MRR	H@10	H@3	H@1	MRR	H@10	H@3	H@1	MRR
LifeKE	<b>0.556</b>	<b>0.408</b>	<b>0.278</b>	<b>0.371</b>	<b>0.563</b>	<b>0.500</b>	<b>0.450</b>	<b>0.487</b>	<b>0.684</b>	<b>0.562</b>	<b>0.415</b>	<b>0.508</b>
w/o neigh.	0.554	0.404	0.270	0.365	0.531	0.477	0.417	0.458	0.677	0.545	0.401	0.495
w/o attn.	0.552	0.405	0.275	0.366	0.553	0.492	0.441	0.479	0.675	0.542	0.411	0.499

图 4 进一步展示了训练过程中验证集上的 H@1 结果. 从图中可以看到: 在数据集 FB15K-237 上, 所提出的注意力机制在模型优化一定轮数后, 可以带来性能的提升. 在 WN18RR 数据集上, 所提出的注意力机制可以加快收敛速度. 这些性能提升主要归功于所提出的注意力机制可以捕捉到和目标链接预测相关的邻居信息, 减少了邻居聚合过程中的噪音和无用信息. 在数据集 WN18RR 上, 所提出的注意力机制给模型带来了相对明显的提升.

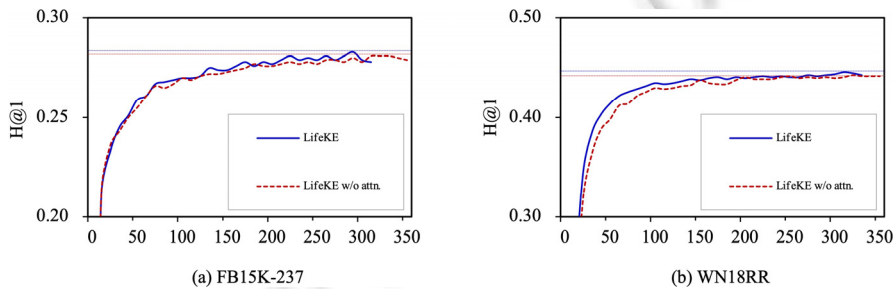


图 4 训练过程中验证集上的 H@1 结果

4.2 多源知识图谱链接预测

本实验基于上述链接预测数据集模拟多源知识图谱终身表示学习的场景, 验证多源知识图谱终身表示学习的可行性以及所提出框架的有效性.

#### 4.2.1 数据介绍

本实验仍以 FB15K-237、WN18RR 和 YAGO3-10 作为多源知识图谱. 为模拟终身表示学习场景, 这 3 个知识图谱依次到来, 所提出的框架依次学习对应的图谱. 本文从 LOD 项目获取这 3 个知识图谱之间的实体对齐数据. 具体来说, 以最流行、数据量也最大的 Wikidata 和 DBpedia 作为中继站, 通过它们到这 3 个数据集的实体对齐数据, 基于传递性规则获得 3 个数据集之间的实体对齐数据, 其实体对齐数量如图 5 所示. 可以看到: FB15K-237 和 YAGO3-10 直接存在较多的实体对齐, 而 WN18RR 到其他两个数据集的实体对齐数量均很少. 需要注意的是, 本文提出的框架依赖于实体对齐进行多源知识迁移.

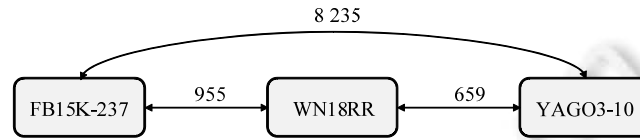


图 5 不同知识图谱之间的实体对齐数量统计

#### 4.2.2 实验设置

实验任务是链接预测. FB15K-237 和 YAGO3-10 是开放域知识图谱, 而 WN18RR 是关于单词上下位关系的词典领域知识图谱. 基于从开放域到特定领域的知识迁移方向, 本实验考虑 4 种以 FB15K-237 或 YAGO3-10 为起点的终身学习顺序(即 4 种多源知识图谱序列): (1) FB15K-237→WN18RR→YAGO3-10; (2) FB15K-237→YAGO3-10→WN18RR; (3) YAGO3-10→FB15K-237→WN18RR; (4) YAGO3-10→WN18RR→FB15K-237. 本文未考虑以 WN18RR 为起点的顺序, 因为预备实验发现, WN18RR 很难为开放域图谱带来知识迁移. 稍后, 本文第 4.2.4 节也对知识图谱可迁移性进行了分析. 此外, 所选择的 4 种顺序也枚举了 3 个知识图谱两两相邻进行迁移的所有情况, 即: FB15K-237→WN18RR; FB15K-237→YAGO3-10; WN18RR→FB15K-237; WN18RR→YAGO3-10; YAGO3-10→FB15K-237; YAGO3-10→WN18RR. 对于每个顺序, 用所提出的框架依次学习每个知识图谱, 然后在相关测试集上进行链接预测性能测试. 由于目前没有相关终身表示学习链接预测的评测方法, 本文以独立训练和联合训练的结果作为基准来评估所提出框架的性能. 独立训练指的是第 4.1 节中的模型训练和测试方法. 每一个知识图谱都独立训练和测试, 这是传统链接预测的任务设定. 联合训练指的是将 3 个知识图谱通过构建并融合两两链接子图, 形成一个大的知识图谱, 然后用所提出的表示学习模型进行训练. 在测试阶段, 分别对 3 个原知识图谱的测试集进行测试. 并且, 只考虑原知识图谱里面的实体作为候选, 来自其他知识图谱的实体不参与计算, 从而使联合训练下的测试候选空间和独立训练下的保持一致, 保证了测试的公平性. 实验结果仍以  $H@k(k=1,3,10)$  和 MRR 作为指标.

#### 4.2.3 主要实验结果

表 4 报告了多源知识图谱终身表示学习的实验结果, 加粗表示相较于独立训练有提升的结果, 斜体表示没有观测到明显提升的结果.

表 4 多源知识图谱终身链接预测性能比较

Metrics	H@10	H@3	H@1	MRR	H@10	H@3	H@1	MRR	H@10	H@3	H@1	MRR
Independent training	FB15K-237				WN18RR				YAGO3-10			
	0.556	0.408	0.278	0.371	0.563	0.500	0.450	0.487	0.684	0.562	0.415	0.508
Joint training	FB15K-237				WN18RR				YAGO3-10			
	<b>0.560</b>	<b>0.408</b>	<b>0.281</b>	<b>0.373</b>	0.525	0.479	0.435	0.466	<b>0.692</b>	<b>0.576</b>	<b>0.415</b>	<b>0.512</b>
Lifelong learning	FB15K-237				WN18RR				YAGO3-10			
	0.556	0.408	0.278	0.371	0.539	0.487	0.447	0.478	<b>0.692</b>	<b>0.565</b>	0.402	0.506
	FB15K-237				YAGO3-10				WN18RR			
	0.556	0.408	0.278	0.371	<b>0.701</b>	<b>0.582</b>	<b>0.430</b>	<b>0.510</b>	0.524	0.481	0.438	0.468
	YAGO3-10				FB15K-237				WN18RR			
	0.684	0.562	0.415	0.508	<b>0.572</b>	<b>0.421</b>	<b>0.289</b>	<b>0.381</b>	0.532	0.482	0.437	0.470
	YAGO3-10				WN18RR				FB15K-237			
	0.684	0.562	0.415	0.508	0.561	0.494	0.440	0.480	0.555	0.405	0.277	0.368

为了深入探究多源知识图谱间表示学习的可迁移性, 本文对实验结果进行了分析与总结, 具体如下.

- (1) 与独立训练相比, 联合训练并未大幅提升效果. 首先, 这是因为 FB15K-237、WN18RR 和 YAGO3-10 这 3 个数据集的大小不平衡, 尤其是 YAGO3-10 的实体和三元组数量远远大于其他两个数据集(见表 2). 这种不平衡会造成训练的不平衡, 导致小图谱训练不充分而大图谱容易过拟合; 其次, 3 个数据集之间实体对齐数量也决定了知识迁移的效果, 其中, FB15K-237 与 YAGO3-10 可以在联合训练中互相增强, 而 WN18RR 因为与其他图谱间的实体对齐较少, 难以获得知识迁移, 表现在没有明显的性能提升. 此外, 由于 WN18RR 是词典数据, 与开放域图谱 FB15K-237 和 YAGO3-10 异构性较强, 这也是其他图谱与 WN18RR 进行联合训练而没有给其带来效果提升的潜在原因;
- (2) 如表 4 中顺序 2 和顺序 3 的结果所示: 在终身表示学习场景下, FB15K-237 可帮助后续 YAGO3-10 的训练, 带来性能提升, 反之亦然. 例如: 所提出模型在 FB15K-237 上独立训练时, H@1 为 0.278; 基于终身表示学习和 YAGO3-10 的基础上(见表中顺序 3), H@1 提高到了 0.289. 同样, YAGO3-10 上的 H@1 结果由独立训练的 0.415 提高到终身学习的 0.430. 这些结果一方面说明了多源知识图谱终身表示学习的可行性, 也验证了所提出框架的有效性. 而在顺序 1 中, 尽管 FB15K-237 和 YAGO3-10 的中间还有 WN18RR, YAGO3-10 上的 H@10 结果 0.692 仍略高于独立训练的 0.684. 这一现象是因为 YAGO3-10 的模型间接获得了来自 FB15K-237 的知识迁移, 且 YAGO3-10 的规模更大, 对知识负迁移的鲁棒性更强. 在顺序 4 中, FB15K-237 的结果非常接近独立训练, 克服了 WN18RR 异构性对迁移的潜在影响. 结合顺序 1 和顺序 4 的结果可以发现: 尽管 WN18RR 不能给后续的知识图谱表示学习带来明显帮助, 但其后的 YAGO3-10 或 FB15K-237 上的结果没有受到较大的负面影响, 表明了本文迁移方法的健壮性;
- (3) 从表 4 中斜体所示结果可以看出, FB15K-237 和 YAGO3-10 这两个图谱对后续的 WN18RR 学习几乎没有帮助. 经深入分析发现, 这主要是由于 WN18RR 和 FB15K-237 与 YAGO3-10 的实体对齐数量较少(见表 1 和图 5 的数据统计). 这说明, 对齐实体是多源知识图谱表示学习在实践中需要关注的重要基础, 对齐实体的数量对终身表示学习有较大影响. 另外, WN18RR 是关于英文单词词义的本体知识库, 而 FB15K-237 和 YAGO3-10 是开放域的知识图谱, 它们的知识异构性较大(见第 4.2.4 节的进一步分析), 可迁移知识较少, 终身学习的性能难以提升.

#### 4.2.4 多源知识图谱可迁移性分析

为进一步探究表 4 实验结果中各知识图谱间迁移效果不同的原因, 本文从图同构的角度入手, 采用 WL 子树核相似度(Weisfeiler-Lehman subtree kernel similarity)<sup>[61]</sup>作为图的拓扑结构相似度指标, 分析 FB15K-237、WN18RR 和 YAGO3-10 这 3 个数据集两两之间的结构相似度, 其结果见表 5. 可以看到, 这 3 个知识图谱间的图同构性都很低. 这是由于 3 个数据集的大小不均, 来源不同, 关系异构性大, 而且图谱间实体对齐数量较少. WL 子树核相似度低也表明了多源知识图谱之间可以直接迁移的知识数量少, 难度大. 相比之下, FB15K-237-YAGO3-10 的 WL 子树核相似度为 0.032 4, 远远高于它们与 WN18RR 的组合. 这也解释了为什么 FB15K-237 和 YAGO3-10 可以在终身学习中互相促进, 而在 WN18RR 上则几乎没有获得收益. 本文认为: 对于结构差异较大的两个图谱, 或实体对齐数量较少的图谱, 可迁移的知识较少, 知识迁移效果会变差. 此外, WL 子树核相似度可以作为多源知识图谱终身表示学习的一个前置验证指标, 以衡量知识迁移的难度以及可以获得的预期收益.

表 5 不同知识图谱间的 WL 子树核相似度比较

	FB15K-237-WN18RR	FB15K-237-YAGO3-10	WN18RR-YAGO3-10
WL subtree kernel sim.	0.006 5	<b>0.032 4</b>	0.001 5

#### 4.2.5 训练时间对比

本节报告独立训练、联合训练和终身学习的每轮训练时间, 结果如图 6 所示. 本实验运行的服务器配置如下.

- CPU: Xeon Gold 6326;

- 内存: 512 GB;
- 显卡: RTX A6000.

可以看到: 联合训练时间开销最大, 因为其需要训练的数据量最大; 独立训练的时间开销最小, 因为模型只需要考虑当前知识图谱的训练; 而本文提出的终身表示学习仅比独立训练增加了少量训练数据(即链接子图和链接实体), 所需时间开销略高于独立训练, 但是远小于联合训练. 由于 YAGO3-10 数据量大, 3 种训练方式所需要的时间都远大于其他两个数据集. 结合表 4 的实验对比, 可以看到, 本文提出的终身表示学习在效果和效率上取得了良好的平衡.

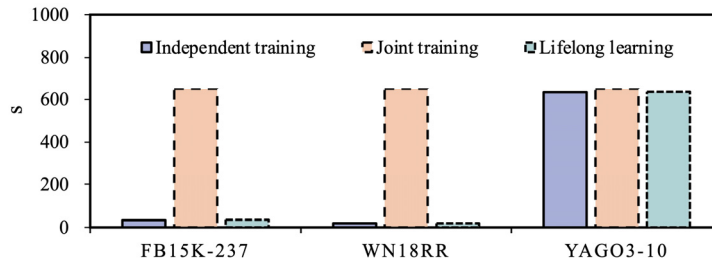


图 6 独立训练、联合训练和终身学习的每轮训练时间对比

## 5 总结与展望

本文研究了多源知识图谱终身表示学习问题, 并提出了一种基于链接实体回放的终身表示学习框架 LifeKE. 框架包含一个基于关系相关性进行链接预测的知识图谱表示学习模型, 其通过实体已有关系和目标关系的相关性聚合实体的邻居信息用于链接预测. 该模型在链接预测基准测试集 FB15K-237 和 WN18RR 上取得了先进的性能. 为了实现知识图谱序列中相邻图谱间的知识迁移和持续学习, 本文所提出的框架通过实体对齐构造新增图谱和已学习图谱之间的链接子图, 并和新增图谱一起用于训练, 且通过子图中链接实体的回放(新表示对齐到旧表示)实现知识迁移. 在链接预测数据集序列上进行的实验证实了多源知识图谱终身表示学习的可行性, 也展示了所提出的框架的效果和效率.

未来工作考虑引入预训练语言模型来增强知识图谱表示学习模型, 并以多模态信息(如文本或者图片等)作为知识迁移的桥梁, 进行更加有效的多源知识图谱终身表示学习. 多源知识图谱终身表示学习对于知识融合、迁移以及应用具有重要的现实意义, 相关代码和数据将开源, 本文提出的方法和实验结果仍有改进空间, 未来工作可以进一步探讨这一方向. 另外, 多源知识图谱中事实的冲突和错误、知识图谱的更新等也是未来研究方向. 例如: 三元组的不一致或实体对齐的噪音等会影响表示学习效果, 可以考虑引入外部资源, 如文本信息或预训练语言模型来辅助表示学习, 或利用对抗训练增强终身表示学习对抗噪音数据的鲁棒性; 事实的新增与删除也要求表示学习模型具有自我更新的能力. 这些问题也是多源知识图谱终身表示学习中值得研究的课题.

## References:

- [1] Ji SX, Pan SR, Cambria E, Marttinen P, Yu PS. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. on Neural Networks and Learning Systems*, 2022, 33(2): 494–514.
- [2] Wang Q, Mao ZD, Wang B, Guo L. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. on Knowledge and Data Engineering*, 2017, 29(12): 2724–2743.
- [3] Liu ZY, Sun MS, Lin YK, Xie RB. Knowledge representation learning: A review. *Journal of Computer Research and Development*, 2016, 53(2): 247–261 (in Chinese with English abstract).
- [4] Chen XL, Chen MH, Fan CJ, Uppunda A, Sun YZ, Zaniolo C. Multilingual knowledge graph completion via ensemble knowledge transfer. In: *Proc. of the EMNLP (Findings)*. 2020. 3227–3238.

- [5] Huang ZJ, Li Z, Jiang HM, Cao TY, Lu HQ, Yin B, Subbian K, Sun YZ, Wang W. Multilingual knowledge graph completion with self-supervised adaptive graph alignment. In: Proc. of the ACL. 2022. 474–485.
- [6] Bordes A, Usunier N, García-Durán A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. In: Proc. of the NIPS. 2013. 2787–2795.
- [7] Wang Z, Zhang JW, Feng JL, Chen Z. Knowledge graph embedding by translating on hyperplanes. In: Proc. of the AAAI. 2014. 1112–1119.
- [8] Lin YK, Liu ZY, Sun MS, Liu Y, Zhu X. Learning entity and relation embeddings for knowledge graph completion. In: Proc. of the AAAI. 2015. 2181–2187.
- [9] Ji GL, He SZ, Xu LH, Liu K, Zhao J. Knowledge graph embedding via dynamic mapping matrix. In: Proc. of the ACL. 2015. 687–696.
- [10] Fang Y, Zhao X, Tan Z, Yang SY, Xiao WD. A revised translation-based method for knowledge graph representation. Journal of Computer Research and Development, 2018, 55(1): 139–150 (in Chinese with English abstract).
- [11] Sun ZQ, Deng ZH, Nie JY, Tang J. RotatE: Knowledge graph embedding by relational rotation in complex space. In: Proc. of the ICLR. 2019.
- [12] Yang BS, Yih TT, He XD, Gao JF, Deng L. Embedding entities and relations for learning and inference in knowledge bases. In: Proc. of the ICLR. 2015.
- [13] Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G. Complex embeddings for simple link prediction. In: Proc. of the ICML. 2016. 2071–2080.
- [14] Kazemi SM, Poole D. Simple embedding for link prediction in knowledge graphs. In: Proc. of the NeurIPS. 2018. 4289–4300.
- [15] Balazevic I, Allen C, Hospedales TM. TuckER: Tensor factorization for knowledge graph completion. In: Proc. of the EMNLP/IJCNLP. 2019. 5184–5193.
- [16] Dettmers T, Minervini P, Stenetorp P, Riedel S. Convolutional 2D knowledge graph embeddings. In: Proc. of the AAAI. 2018. 1811–1818.
- [17] Guo LB, Sun ZQ, Hu W. Learning to exploit long-term relational dependencies in knowledge graphs. In: Proc. of the ICML. 2019. 2505–2514.
- [18] Vashishth S, Sanyal S, Nitin V, Talukdar PP. Composition-based multi-relational graph convolutional networks. In: Proc. of the ICLR. 2020.
- [19] Wang S, Wei XK, dos Santos CN, Wang ZG, Nallapati R, Arnold AO, Xiang B, Yu PS, Cruz IF. Mixed-curvature multi-relational graph neural network for knowledge graph completion. In: Proc. of the WWW. 2021. 1761–1771.
- [20] Wang Q, Huang PP, Wang HF, Dai ST, Jiang WB, Liu J, Lyu YJ, Zhu Y, Wu H. CoKE: Contextualized knowledge graph embedding. CoRR abs/1911.02168, 2019.
- [21] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. In: Proc. of the NIPS. 2017. 5998–6008.
- [22] Zhang NY, Xie X, Chen X, Deng SM, Ye HB, Chen HJ. Knowledge collaborative fine-tuning for low-resource knowledge graph completion. Ruan Jian Xue Bao/Journal of Software, 2022, 33(10): 3531–3545 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6628.htm> [doi: 10.13328/j.cnki.jos.006628]
- [23] Zhang TC, Tian X, Sun XH, Yu MH, Sun YH, Yu G. Overview on knowledge graph embedding technology research. Ruan Jian Xue Bao/Journal of Software, 2023, 34(1): 277–311 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6429.htm> [doi: 10.13328/j.cnki.jos.006429]
- [24] Chen MH, Tian YT, Yang MH, Zaniolo C. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In: Proc. of the IJCAI. 2017. 1511–1517.
- [25] Sun ZQ, Hu W, Li CK. Cross-lingual entity alignment via joint attribute-preserving embedding. In: Proc. of the ISWC. 2017. 628–644.
- [26] Sun ZQ, Hu W, Zhang QH, Qu YZ. Bootstrapping entity alignment with knowledge graph embedding. In: Proc. of the IJCAI. 2018. 4396–4402.

- [27] Wang ZC, Lv QS, Lan XH, Zhang Y. Cross-lingual knowledge graph alignment via graph convolutional networks. In: Proc. of the EMNLP. 2018. 349–357.
- [28] Trisedya BD, Qi JZ, Zhang R. Entity alignment between knowledge graphs using attribute embeddings. In: Proc. of the AAAI. 2019. 297–304.
- [29] Yang HW, Zou YY, Shi P, Lu W, Lin J, Sun X. Aligning cross-lingual entities with multi-aspect information. In: Proc. of the EMNLP/IJCNLP. 2019. 4430–4440.
- [30] Zeng WX, Zhao X, Tang JY, Lin XM, Groth P. Reinforcement learning-based collective entity alignment with adaptive features. ACM Trans. on Information Systems, 2021, 39(3): 26:1–26:31.
- [31] Tang XB, Zhang J, Chen B, Yang Y, Chen H, Li CP. BERT-INT: A BERT-based interaction model for knowledge graph alignment. In: Proc. of the IJCAI. 2020. 3174–3180.
- [32] Ge CC, Liu XZ, Chen L, Zheng BH, Gao YJ. Make it easy: An effective end-to-end entity alignment framework. In: Proc. of the SIGIR. 2021. 777–786.
- [33] Chen LY, Li Z, Wang YJ, Xu T, Wang ZF, Chen EH. MMEA: Entity alignment for multi-modal knowledge graph. In: Proc. of the KSEM. 2020. 134–147.
- [34] Liu FY, Chen MH, Roth D, Collier N. Visual pivoting for (unsupervised) entity alignment. In: Proc. of the AAAI. 2021. 4257–4266.
- [35] Chen MH, Tian YT, Chang KW, Skiena S, Zaniolo C. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment. In: Proc. of the IJCAI. 2018. 3998–4004.
- [36] Zeng WX, Zhao X, Tang JY, Tan Z, Wang W. Iterative entity alignment via re-ranking. Journal of Computer Research and Development, 2020, 57(7): 1460–1471 (in Chinese with English abstract).
- [37] Sun ZQ, Zhang QH, Hu W, Wang CM, Chen MH, Akrami F, Li CK. A benchmarking study of embedding-based entity alignment for knowledge graphs. Proc. of the VLDB Endowment, 2020, 13(11): 2326–2340.
- [38] Zhao X, Zeng WX, Tang JY, Wang W, Suchanek FM. An experimental study of state-of-the-art entity alignment approaches. IEEE Trans. on Knowledge and Data Engineering, 2022, 34(6): 2610–2625.
- [39] Han YN, Liu JW, Luo XL. Research progress of continual learning. Journal of Computer Research and Development, 2022, 59(6): 1213–1239 (in Chinese with English abstract).
- [40] Kirkpatrick J, Pascanu R, Rabinowitz NC, Veness J, Desjardins G, Rusu AA, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, Hassabis D, Clopath C, Kumaran D, Hadsell R. Overcoming catastrophic forgetting in neural networks. PNAS, 2017, 114(13): 3521–3526.
- [41] Lopez-Paz D, Ranzato MA. Gradient episodic memory for continual learning. In: Proc. of the NIPS. 2017. 6467–6476.
- [42] Yoon JH, Yang EH, Lee JT, Hwang SJ. Lifelong learning with dynamically expandable networks. In: Proc. of the ICLR. 2018.
- [43] Daruna AA, Gupta M, Sridharan M, Chernova S. Continual learning of knowledge graph embeddings. IEEE Robotics and Automation Letters, 2021, 6(2): 1128–1135.
- [44] Kou XY, Lin YK, Liu SB, Li P, Zhou J, Zhang Y. Disentangle-based continual graph representation learning. In: Proc. of the EMNLP. 2020. 2961–2972.
- [45] Dong LJ, Zhao DY, Zhang XA, Li XC, Kang XJ, Yao H. Anchors-based incremental embedding for growing knowledge graphs. IEEE Trans. on Knowledge and Data Engineering, 2021.
- [46] Teru KK, Denis EG, Hamilton WL. Inductive relation prediction by subgraph reasoning. In: Proc. of the ICML. 2020. 9448–9457.
- [47] Hamaguchi T, Oiwa H, Shimbo M, Matsumoto Y. Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach. In: Proc. of the IJCAI. 2017. 1802–1808.
- [48] Wang PF, Han JL, Li CL, Pan R. Logic attention based neighborhood aggregation for inductive knowledge graph embedding. In: Proc. of the AAAI. 2019. 7152–7159.
- [49] Hu WH, Liu BW, Gomes J, Zitnik M, Liang P, Pande VS, Leskovec J. Strategies for pre-training graph neural networks. In: Proc. of the ICLR. 2020.
- [50] Fang Y, Zhao X, Chen YF, Xiao WD, de Rijke M. PF-HIN: Pre-training for heterogeneous information networks. IEEE Trans. on Knowledge and Data Engineering, 2022.



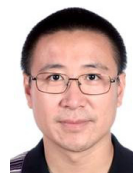
- [51] Bollacker KD, Evans C, Paritosh PK, Sturge T, Taylor J. Freebase: A collaboratively created graph database for structuring human knowledge. In: Proc. of the SIGMOD Conf. 2008. 1247–1250.
- [52] Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives ZG. DBpedia: A nucleus for a Web of open data. In: Proc. of the ISWC/ASWC. 2007. 722–735.
- [53] Vrandečić D, Krötzsch M. Wikidata: A free collaborative knowledgebase. Communications of the ACM, 2014, 57(10): 78–85.
- [54] Mahdisoltani F, Biega J, Suchanek FM. YAGO3: A knowledge base from multilingual Wikipedias. In: Proc. of the CIDR. 2015.
- [55] Zhang ZQ, Wang J, Ye JP, Wu F. Rethinking graph convolutional networks in knowledge graph completion. In: Proc. of the WWW. 2022. 798–807.
- [56] Yan YC, Liu LH, Ban YK, Jing BY, Tong HH. Dynamic knowledge graph alignment. In: Proc. of the AAAI. 2021. 4564–4572.
- [57] Toutanova K, Chen DQ. Observed versus latent features for knowledge base and text inference. In: Proc. of the CVSC. 2015. 57–66.
- [58] Miller GA. WordNet: A lexical database for English. Communications of the ACM, 1995, 38(11): 39–41.
- [59] Glorot X, Bengio Y. Understanding the difficulty of training deep feed forward neural networks. In: Proc. of the AISTATS. 2010. 249–256.
- [60] Kingma DP, Ba J. ADAM: A method for stochastic optimization. In: Proc. of the ICLR. 2015.
- [61] Shervashidze N, Schweitzer P, van Leeuwen EJ, Mehlhorn K, Borgwardt KM. Weisfeiler-Lehman graph kernels. Journal of Machine Learning Research, 2011, 12: 2539–2561.

#### 附中文参考文献:

- [3] 刘知远, 孙茂松, 林衍凯, 谢若冰. 知识表示学习研究进展. 计算机研究与发展, 2016, 53(2): 247–261.
- [10] 方阳, 赵翔, 谭真, 杨世宇, 肖卫东. 一种改进的基于翻译的知识图谱表示方法. 计算机研究与发展, 2018, 55(1): 139–150.
- [22] 张宁豫, 谢辛, 陈想, 邓淑敏, 叶宏彬, 陈华钧. 基于知识协同微调的低资源知识图谱补全方法. 软件学报, 2022, 33(10): 3531–3545. <http://www.jos.org.cn/1000-9825/6628.htm> [doi: 10.13328/j.cnki.jos.006628]
- [23] 张天成, 田雪, 孙相会, 于明鹤, 孙艳红, 于戈. 知识图谱嵌入技术研究综述. 软件学报, 2023, 34(1): 277–311. <http://www.jos.org.cn/1000-9825/6429.htm> [doi: 10.13328/j.cnki.jos.006429]
- [36] 曾维新, 赵翔, 唐九阳, 谭真, 王炜. 基于重排序的迭代式实体对齐. 计算机研究与发展, 2020, 57(7): 1460–1471.
- [39] 韩亚楠, 刘建伟, 罗雄麟. 连续学习研究进展. 计算机研究与发展, 2022, 59(6): 1213–1239.



孙泽群(1992—), 男, 博士生, CCF 学生会员, 主要研究领域为知识图谱表示学习.



胡伟(1982—), 男, 博士, 副教授, 博士生导师, CCF 高级会员, 主要研究领域为知识图谱, 数据集成, 智能软件.



崔员宁(1996—), 男, 博士生, 主要研究领域为知识图谱表示学习.